

1/39

FIG. 1A

2736 bp

```

1  ATGGCGCCGA GCTCCCCGTC ACCCGCCGCG CCTACACGGG TCTCTGGGCG
51  GAAGCGCGCC GCCAAGGCGG AGGAGATCCA CCAGAACAAG GAGGAGGAGG
101 AGGAGGTGCG GGGCGCGTCT TCCGCCAAGC GCAGCCGCAA GCGCGCATCT
151 TCCGGGAAGA AGCCCAAGTC GCCCCCAAG CAGGCCAAGC CGGGGAGGAA
201 GAAGAAGGGG GATGCCGAGA TGAAGGAGCC CGTGGAGGAC GACGTGTGCG
251 CCGAGGAGCC CGACGAGGAG GAGTTGGCCA TGGGCGAGGA GGAGGCCGAG
301 GAGCAGGCCA TGCAGGAGGA GGTGGTTGCG GTCGCGGCGG GGTCAACCGG
351 GAAGAAGAGG GTGGGGAGAA GGAACGCCGC CGCCGCGCGT GCGGACCACG
401 AGCCGGAGTT CATCGGCAGC CCTGTTGCCG CGGACGAGGC GCGC/GCAAC
451 TGGCCCAAGC GCTACGGCCG CAGCAC/TGCC GCAAAGAAAC CGGATGAGGA
501 GGAAGAGCTC AAGGCCAGAT GTCACTACCG GAGCGCTAAG GTGGACAACG
551 TCGTCTACTG CCTCGGGGAT GACGTCATG TCAAGGCTGG AGAAAACGAG
601 GCAGATTACA TTGGCCGCAT TACTGAATTT TTTGAGGGGA CTGACCAGTG
651 TCACTATTTT ACTTGCCGTT GGTTCCTCCG AGCAGAGGAC ACGGTTATCA
701 ATTCTTTGGT GTCCATAAGT GTGGATGGCC ACAAGCATGA CCTAGACGT
751 GTTTTTCTTT CTGAGGAAAA GAACGACAAT GTGCTTGATT GCATTATCTC
801 CAAGGTCAAG ATAGTCCATG TTGATCCAAA TATGGATCCA AAAGCCAAAG
851 CTCAGCTGAT AGAGAGTTGC GACCTATACT ATGACATGTC TTA CTCTGTT
901 GCATATTCTA CATTTGCTAA TATCTCGTCT GAAAATGGGC AGTCAGGCAG
951 TGATACCGCT TCGGGTATTT CTTCTGATGA TGTGGATCTG GAGACGTCAT
1001 CTAGTATGCC AACGAGGACA GCAACCTTTC TTGATCTGTA TTCTGGCTGT
1051 GGGGGCATGT CTA CTGGTCT TTGCTTGGGT GCAGCTCTTT CTGGCTTGAA
1101 ACTTGAAACT CGATGGGCTG TTGATTTCAA CAGTTTTGCG TGCCAAAGTT
1151 TAAAATATAA TCATCCACAG ACTGAGGTGC GAAATGAGAA AGCCGATGAG
1201 TTTCTTGCCC TCCTTAAGGA ATGGGCAGTT CTATGCAAAA AATATGTCCA
1251 AGATGTGGAT TCAAATTTAG CAAGCTCAGA GGATCAAGCG GATGAAGACA
1301 GCCCTCTTGA CAAGGACGAA TTTGTGTAG AGAAGCTTGT CGGGATATGT
1351 TATGGTGGCA GTGACAGGGA AAATGGCATC TATTTTAAGG TCCAGTGGGA
1401 AGGATACGGC CCTGAGGAGG ATACATGGGA ACCGATTGAT AACTTGAGTG
1451 ACTGCCCGCA GAAATTTAGA GAATTTGTAC AAGAAGGGCA CAAAAGAAAG
1501 ATTCTCCAC TGCCTGGTGA TGTGTATGTC ATTTGTGGAG GCCCACCATG
1551 CCAAGGTATC AGTGGGTTTA ATCGGTACAG AAACCGTGAT GAGCCACTCA
1601 AAGATGAGAA AAACAAACAA ATGGTGACTT TCATGGATAT TGTGGCGTAC
1651 TTGAAGCCCA AGTATGTTCT CATGGAAAAT GTGGTGGACA TACTCAAATT
1701 TGCGGATSGT TACCTAGGAA AATATGCTTT GAGCTGCCTT GTTGCTATGA
1751 AGTACCAAGC GCGGCTTGGG ATGATCGTGG CTGGTTGCTA TGGTCTGCCA
1801 CAGTTCAGGA TGCCTGTGTT CCTCTGGGGT GCTCTTTCTT CCATGGTGCT
1851 CCTAAGTAT CCTCTGCCCA CCTATGATGT TGTAGTACGT GGAGGAGCCC
1901 CTAATGCGTT TTGCAATGT ATGGTTGCAT ATGACGAGAC ACAAAAACCA
1951 TCCCTGAAAA AAGCCTTGCT TCTTGGCGAT GCAATTTAG ATTTACCAAA
2001 GGTTCAAAAT CACCAGCCTA ACGATGTGAT GGAGTATGGT GGTTCGCCCA
2051 AGACCGAATT CCAGCGCTAC ATTCGACTCA GTCGTAAAGA CATGTTGGAT
2101 TGGTCCTTCG GTGAGGGGGC TGGTCCAGAT GAAGGCAAGC TCTTGGATCA
2151 CCAGCCTTTA CCGCTTAACA ACGATGATTA TGAGCGGGTT CAACAGATTG
2201 CTGTCAAGAA GGGAGCCAAC TTCCGCGACC TAAAGGGCGT GAGGGTTGGA
2251 GCAAACAATA TTGTGAGTG GGATCCAGAA ATCGAGCGTG TGAAACTTTC
2301 ATCTGGGAAA CCACTGCTTC CTGACTATGC AATGTCATTC ATCAAGGGCA
2351 AATCACTCAA GCCSTTTGGG CGCCTGTGGT GGGACGAGAC AGTTCCTACA
2401 GTTGTAAACA GAGCAGAGCC TCACAACCAG GTTATAATTC ATCCGACTCA
2451 AGCAAGGGTC CTCACTATCC GGGAGAACGC AAGGTTACAG GCTTCCCGG
2501 ATTACTACCG ATTGTTTGGC CCGATCAAGG AGAAGTACAT TCAAGTCGGG
2551 AACGCACTGG CTGTCCCTGT TGCCCGGGCA CTGGGCTACT GTCTGGGGCA
2601 AGCCTACCTG GGTGAATCTG AGGGGAGTGA CCTCTGTAC CAGCTGCCTC
2651 CAAGTTTAC CTCTCTTGGG GAGCGCACTG CGGGGAGGCG GAGGCGCTCT
2701 CCTGTTGGCA CCGCTGCAGG GGAGGTAGTT GAGCAG

```

2/39
FIG. 1B

1 AGAGCAGCAG CAGCTACCGC AGCCCTGCC ATGGCGCCGA GCTCCCCGTC
 51 ACCCGCCGCG CCTACACGCG TCTCTGGGCG GAAGCGCGCC GCCAAGGCCG
 101 AGGAGATCCA CCAGAACAAG GAGGAGGAGG AGGAGGTCGC GCGCGCGTCC
 151 TCCGCCAAGC GCAGCCGCAA GCGCGCATCT TCCGGGAAGA AGCCCAAGTC
 201 GCCCCCAAG CAGGCCAAGC CGGGGAGGAA GAAGAAGGGG GATGCCGAGA
 251 TGAAGGAGCC CGTGGAGGAC GACGTGTGCG CCGAGGAGCC CGACGAGGAG
 301 GAGTTGGCCA TGGGCGAGGA GGAGGCCGAG GAGCAGGCCA TGCAGGAGGA
 351 GGTGGTTGCG GTCGCGGCGG GGTCAACCCG GAAGAAGAGG GTGGGGAGAA
 401 GGAACGCCGC CGCCGCGGCT GCGGACCACG AGCCGGAGTT CATCGGCAGC
 451 CCTGTTGCCG CGGACGAGGC CGCAGCAAC TGGCCCAAGC GCTACGGCCG
 501 CAGCACTGCC GCAAAGAAAC CGGATGAGGA GGAAGAGCTC AAGGCCAGAT
 551 GTCACTACCG GAGCGCTAAG GTGGACAACG TCCTCTACTG CCTCGGGGAT
 601 GACGCTCTATG TCAAGGCTGG AGAAAACGAG GCAGATTACA TTGGCCGCAT
 651 TACTGAATTT TTTGAGGGGA CTGACCAAGT TCACCTATTTT ACTTGCCGTT
 701 GGTTCCTCCG AGCAGAGGAC ACGGTTATCA ATTCTTTGGT GTCCATAAGT
 751 GTGGATGGCC ACAAGCATGA CCCTAGACGT GTTTTCTTT CTGAGGAAAA
 801 GAACGACAAT GTGCTTGATT GCATTATCTC CAAGGTCAAG ATAGTCCATG
 851 TTGATCCAAA TATGGATCCA AAAGCCAAGG CTCAGCTGAT AGAGAGTTGC
 901 GACCTATACT ATGACATGTC TTACTCTGTT GCATATTCTA CATTTGCTAA
 951 TATCTCGTCT GAAAATGGGC AGTCAGGCAG TGATACCGCT TCGGGTATTT
 1001 CTTCTGATGA TGTGGATCTG GAGACGTCAT CTAGTATGCC AACGAGGACA
 1051 GCAACCCCTC TTGATCTGTA TTCTGGCTGT GGGGGCATGT CTACTGGTCT
 1101 TTGCTTGGGT GCAGCTCTTT CTGGCTTGA ACTTGAAACT CGATGGGCTG
 1151 TTGATTCAA CAGTTTTGCG TGCCAAAGTT TAAAATATAA TCATCCACAG
 1201 ACTGAGGTGC GAAATGAGAA AGCCGATGAG TTTCTTGCCC TCCTTAAGGA
 1251 ATGGGCAGTT CTATGCAAAA AATATGTCCA AGATGTGGAT TCAAATTTAG
 1301 CAAGCTCAGA GGATCAAGCG GATGAAGACA GCGCTCTTGA CAAGGACGAA
 1351 TTTGTTGTAG AGAAGCTTGT CGGGATATGT TATGGTGGCA GTGACAGGGA
 1401 AAATGGCATC TATTTTAAGG TCCAGTGGGA AGGATACGGC CCTGAGGAGG
 1451 ATACATGGGA ACCGATTGAT AACTTGAGTG ACTGCCCGCA GAAAATTAGA
 1501 GAATTTGTAC AAGAAGGGCA CAAAAGAAAG ATTCTCCAC TGCCTGGTGA
 1551 TGTGATGTC ATTTGTGGAG GCCCACCATG CCAAGGTATC AGTGGGTTTA
 1601 ATCGGTACAG AAACCGTGT GAGCCACTCA AAGATGAGAA AAACAAACAA
 1651 ATGGTGACTT TCATGGATAT TGTGGCGTAC TTGAAGCCCA AGTATGTTCT
 1701 CATGGAAAAT GTGGTGGACA TACTCAAATT TGCGGATGGT TACCTAGGAA
 1751 AATATGCTTT GAGCTGCCTT GTTGCTATGA AGTACCAAGC GCGGCTTGGA
 1801 ATGATGGTGG CTGGTTGCTA TGGTCTGCCA CAGTTCAGGA TGCGTGTGTT
 1851 CCTCTGGGGT GCTCTTTCTT CCATGGTGCT CCTAAGTAT CCTCTGCCCA
 1901 CCTATGATGT TGTAGTACGT GGAGGAGCCC CTAATGCCTT TTCGCAATGT
 1951 ATGGTTGCAT ATGACGAGAC ACAAAAACCA TCCCTGAAAA AAGCCTTGCT
 2001 TCTTGGCGAT GCAATTTTCA ATTTACCAAA GGTTCAAAAT CACCAGCCTA
 2051 ACGATGTGAT GGAGTATGGT GGTTCGCCCA AGACCGAATT CCAGCGCTAC
 2101 ATTCGACTCA GTCGTAAAGA CATGTTGGAT TGGTCTTCG GTGAGGGGCG
 2151 TGGTCCAGAT GAAGGCAAGC TCTTGATCA CCAGCCTTTA CGGCTTAACA
 2201 ACGATGATTA TGAGCGGGTT CAACAGATTG CTGTCAAGAA GGGAGCCAAC
 2251 TTCCGCGACC TAAAGGGCGT GAGGGTTGGA GCAAACAATA TTGTTGAGTG
 2301 GGATCCAGAA ATCGAGCGTG TGAAACTTTC ATCTGGGAAA CCACTGGTTC
 2351 CTGACTATGC AATGTCATTG ATCAAGGGCA AATCACTCAA GCGGTTTGGG
 2401 CGCCTGTGGT GGGACGAGAC AGTTCTTACA GTTGTAAACCA GAGCAGAGCC
 2451 TCACAACCAAG GTTATAATTG ATCCGACTCA AGCAAGGGTC CTCACATTCC
 2501 GGGAGAACGC AAGGTTACAG GGCTTCCCCG ATTACTACCG ATTGTTTGGC
 2551 CCGATCAAGG AGAAGTACAT TCAAGTCGGG AACGCAAGTG CTGTCCCTGT
 2601 TGCCCGGGCA CTGGGCTACT GTCTGGGCGA AGCCTACCTG GGTGAATCTG
 2651 AGGGGAGTGA CCTCTGTAC CAGCTGCCTC CAAGTTTAC CTCTGTGGA
 2701 GGAGCAGCTG CGGGCGAGGC GAGGGCCTCT CCTGTTGGCA CCCCTGCAAG
 2751 GGAGGTAGTT GAGCAGTAA AGGATGACAG ATCTGAGCTG AGCTGG

3/39

FIG. 2A

912 amino acids

1	MAPSSPSPAA	PTRVSGRKRA	AKAEETIQNK	EEEEVAAAS	SAKRSRKAAS
51	SGKKPKSPPK	QAKPGRKKKG	DAEMKEPVED	DVCAEEPDEE	ELAMGEEAE
101	EQAMQEEVVA	VAAGSPGKKR	VGRRNAAAAA	GDHEPEFIGS	PVAADERSN
151	WPKRYGRSTA	AKKPDEEEEL	KARCHYRSK	VDNVVYCLGD	DVYVKAGENE
201	ADYIGRITF	FEGTDQCHYF	TCRWFFRAED	TVINSLVSIS	VDGHKHDPRR
251	VFLSEEKNDN	VLDCIISKVK	IVHVDPNMDP	KAKAQLIESC	DLYYDMSYSV
301	AYSTFANISS	ENGQSGSDTA	SGISSDDVDL	ETSSSMPTRT	ATLLDLYSGC
351	GGMSTGLCLG	AALSGLKLET	RWAVDFNSFA	CQSLKYNHPQ	TEVRNEKADE
401	FLALLKEWAV	LCKKYVQDVD	SNLASEDQA	DEDSPLDKDE	FVVEKLVGIC
451	YGGSDRENGI	YFKVQWEGYG	PEEOTWEPID	NLSDCPQKIR	EFVQEGHKRK
501	ILPLPGDQDV	ICGGPPCQGI	SGFNRYRNRD	EPLKDEKNKQ	MVTFMDIVAY
551	LKPKYVLMEN	VVDILKFADG	YLGKYALSCL	VAMKYQARLG	MMVAGCYGLP
601	QFRMRVFLWG	ALSSMVLPHY	PLPTYDVVVR	GGAPNAFSQC	MVAYDETQKE
651	SLKKALLLGD	AISDLPKVQN	HQPNDVMEYG	GSPKTEFQRY	IRLSRKDMLD
701	WSFGEGAGPD	EGKLLDHQPL	RLNNDYERV	QQIPVKKGAN	FRDLKGVRVG
751	ANNIVEWDPE	IERVKLSSGK	PLVPDYAMSF	IKGKSLKPFQ	RLWWDVETVPT
801	VVTRAEPHNQ	VIIHPTQARV	LTIRENARLQ	GFPDYRLFG	PIKEYIQVG
851	NAVAVPVARA	LGYCLGQAYL	GESEGSPLY	QLPPSFTSVG	GRTAGQARAS
901	PVGTPAGEVV	EQ			

4/39

FIG. 2B

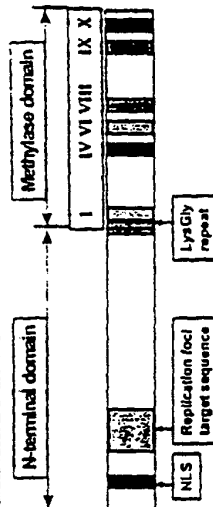

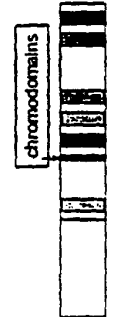
RAAATAAPAPAPSSPSPAAPTRVSGRRAAKAEETHQNKEEEEVAAAS
 SAKRSRKAASSGKKPKSPKQAKPGRKKKGDAEMKEPVEDDVCAEEPDEE
 ELAMGEEEAEEQAMQEEVVAAGSPGKKRVGRNAAAAAGDHEPEFTGS
 PVAADARSNWPKRYGRSTAACKPDDEEELKARCHYRSKVDNVVYCLGD
 DVYVKAGENEADYIGRITEFFEGTDQCHYFTCRWFFRAEDTVINSLVSI
 VDGHKHDFAFVFLSEKNDNVLDCTISKVKIVHVDPNMOPKAKAQLIESC
 DLYYDMSYSVAYSTFANISSENGQSGSDTASGISDDVDLETSSMPTRT
 ATLLDLYSGCGGMYSTGLCLGAALSGLKLETRWAVDFNSFACQSLKYNHPQ
 TEVRNEKADEFTALLKEWAVLCKKYVQDVDSNLASSEDQADEDSPLDKDE
 FVVEKLVGICYGGSDRENGIYFKVQWEGYGPEEDTWEPIDNLSDCPQKIR
 EFVQEGHKRKILPLPGDVDVICGGPPCQGISGFNRNRYNRDEPLKDEKNKQ
 MVTFMDIVAYLKPXYVLMENVVDILKFADGYLGKYALSCLVAMKYQARLG
 MMVAGCYGLPQFRMRVFLWGALSSMVLPKYPLPTYDVVVRGGAPNAPSQC
 MVAYDETQKPSLKALLLGDALSDLPKVQNHQPNVMEYGGSPKTEFQRY
 IRLSRKMDLWSFGGAGPDEGKLLDHQPLRLNNDYERVQQIPVKKGAN
 FRDLKGVVRVGANNIVEWDPEIERVKLSSGKPLVPDYAMSFYKGSCLKPFG
 RLWWDDETVPVTVTRAEPHNQVIHPTQARVLTIRENARLQGFDPYYRLF
 PIKEYIQVGNVAVPVARALGYCLGQAYLGESEGSPLYQLPPSFTSVG
 GRTAGQARASPVGTPAGEVVEQ*KDDRSELSW

5/39

FIG. 3

Primer	Sequence 5' - 3'
1F	TGGTTGCTATGGTCTGCCACAGTTCAG
1R	CCAGCTCAGCTCAGATCTGTCATCCTTT
Seq2FN	CGAAAGCTAATCTACACAAACAGC
Seq2RN	GATCCTCTGAGCTTGCTAAATTTG
3R	CTCATCTTGGAGTGGCTCATCAC
S3F	GAGCACATGAGGGAGAGTGTTG
S3R	TCTCTAATTTTCTGCGGGCAG
4F	CCTCTGCCACCTATGATGTTGTA
5F	TAAAGGGCGTGAGGGTTGGA
7F	TCACATTTGTCATGGCAGGTTATC
8eF	CTGAGGAAAAGAACGACAATGTGC
8eR	GCAATCAAGCACATTGTCGTTCTTTTCCTC
9eF	GAAGAAGAGGGTGCGGAGAAGGAACG
9eR	TTCTTTGCGGCAGTGCTGCG
11iF	GTATTGAATTGATTCTCAACTAGTGCAC
11iR	CAGGCTCAACGGCGATG
12iF	TATGCTTCATCACATAGACCCAAGTC
12iR	GATAGACCTAATGCCAAATGAGATTAAG
13iF	GCGATCTTCAGTCTCCACCATC
13iR	GAAGACGTGCCTCCATGTTTCATC
14F	GTTGGTTCTTCCGAGCAGAGG
14R	GACTGCCACATATCTTATTAATCGC
15F	GCATGTGTCAGCAATTGCTTACATTC
15R	CCTCTGCTCGGAAGAACCAAC
16F	CTGTTCCGAGATTCATGCATGATG
16R	GGAGAACAGAATGGTTGATTCAATGG
17F	GCACTTCACTCTCCTGGCAAACC
17R	CGGTACGCTGCTGCTGCTCTC
18F	CCATAGCATCTCACATATCGCAAGG
18R	GGAAAGAAGGCAGTTAGTTGTAAATGGG
MU	AGAGAAGCCAACGCCAWCGCCTCYATTTGTC
RaceRT	CTACAACATCATAGTTGGGCAGAGG
AP2 marathon	ACTCACTATAGGGCTCGAGCGGC
T7	TAATACGACTCACTATAGGG
Sp6	GATTTAGGTGACACTATAG
M13F	GTTTCCCAGTCACGAC
M13R	CAGGAAACAGCTATGAC

FIG. 4

Gene Name	Organism	Function	General Structure
<i>DNMT1/Dnmt1</i>	human/mouse	maintenance	
<i>ME1</i>	<i>Arabidopsis</i>	maintenance	
<i>Zmct1</i>	maize	putative maintenance	
<i>DNMT3/Dnmt3</i>	human/mouse	<i>de novo</i>	
<i>Zmct3</i>	maize	putative <i>de novo</i>	
<i>DRM</i>	<i>Arabidopsis</i>	putative <i>de novo</i>	
<i>CMT1</i>	<i>Arabidopsis</i>	undetermined (putative CpNpG)	
<i>Zmct2α</i>	maize	CpNpG (maintenance and/or <i>de novo</i>)	

7/39

Figure 5

HhaI MseI
 tacatcaataaaataagggcgcccaacgcaattgtcccttGtttttttctaacttaagttcaagcggcaatgtcg base pairs
 atgtagttattttattcccccggttgcttaacaggggaacaaaaagattgaatttcaagttcccggttacagc 1 to 75
 MseI MseI
 catctgatgtatgaatatcaatttgaagtactaaacatagctttaaattccaaaatttaattacacaacaaagact base pairs
 gtagactacatacttatagttaaacattcatgatttgtatcagaattaagggttttaattaatgtgtgtttctga 76 to 150
 MseI HinfI
 aaattgtaagcaaacctttcaagtctaatttaattcataattacaattgtattgtaacatcatgttacgaatca base pairs
 tttacattcggttggaaggtcagattaattaagtattaatgtttacaataacattgtagtacaatggcttagt 151 to 225
 ScrFI EcoRII MseI
 taaactaaccaggttcccatgtgtaattagttttataattatattatatttaattttgtaactaattgatgtga base pairs
 atttgattggtccaaggtacacatttaatacaaatatttaataacataaattataaacattgattaactacact 226 to 300
 BstNI
 MseI MseI MseI
 cagtactaaaataagcctcttaagccaaaaatccacataatttagattttaaatttgaaaacagacgtatcgg base pairs
 gtcatgattttaattcggagaattcgggttttttaggtgtataaaattctaaattttaaaactttgtctgcataagcc 301 to 375
 HaeIII
 ctagaagagccctgtcactgtcagctaattcaattacaagaagtgcccatatactagttccatcaccagtcagtag base pairs
 gatcttctcgggacagtgacagtcgattagttaatgtcttccacgggtatgatcaaggtagtggtcaggtcatc 376 to 450
 PvuII HaeIII HaeIII HhaI HhaI HpaII
 tccaccacccccacctacagctgggtcatctggcacgggtggagggggccaaagcggccaaagcggcgcaacttcc base pairs
 aggtggtgggggtgggatgtcgaccagtagacgtgcccacctccccgggtgcccgttttccggcgcggtgaagg 451 to 525
 MspI
 HinfI PstI Eco109I
 ggcgggcaccctCgcgagtcgcggtgacagcgaaatttcaaatccataccctcccgctgcagacgggccccac base pairs
 ccgcccgtgggagcgccctcagcgccactgtcgctttaaagttaggtatgggagggcgacgtctgcccggggtg 526 to 600
 HaeIII
 TaqI
 gccgtcaaaatttggacgtccccgtccctcgatcttttgggtttcgttttccagttccccaccctctcttccac base pairs
 cggcagttttaaactgcgagggcgagggttagaataacccaaagcaaaaggggtcaagggtgggagagaagggtg 601 to 675
 Sau3AI
 Sau3AI TaqI TaqI HinfI
 cctgccctgtttccagatttgaccgatcccccttcgattcgatttctacacccacgggtgtccagactccagagcac base pairs
 ggacgggacaaaaggtctaaactggctagggggaagctaagctaagatgtgggtgccacaggtctgaggtctctgtg 676 to 750
 HinfI
 ScrFI
 17F EcoRII
 tcaactcctcctggcgaaccccttctgcttccccaacctagagagcagcagcgtaccgcagccctgcatggc base pairs
 agtgagaggaccgtttgggaaagcagaaggttgggatctctctcgtcgtcgaaggcgtcggggacgggtaccg 751 to 825
 BstNI 17R
 HhaI SacI HhaI HhaI HaeIII Sau3AI
 gccgagctccccgtcacccgcgcgcctacacgcgtctctggcggaagcgcgcgcgaaggccgaggagatcca base pairs
 cggctcgaggggcagtgggcggcgcggatgtgcgcagagaccgccttcgcgcgcgggttccggtcctcttaggt 826 to 900
 HhaI HpaII
 ccagaacaaggaggaggaggagggtcgcgggcggtcctccgccaagcgcagccgcaaggcggcatcttccgg base pairs
 ggtctgttctcctcctcctcctccagcgccgcgcagggcggttcgcgtcggcggttccgctgagaaggcc 901 to 975
 MspI
 MspI HaeIII ScrFI
 gaagaagcccaagtcgcccccaagcagggccaagccggggagggaagaaggggggtatgccgagatgaaggagcc base pairs
 ctctctcgggttcagcgggggttcgtccggttcggccctcctctcttccccctacggctctacttctctcgg 976 to 1050
 HpaII

8/39
FIG. 5

HhaI Continued HaeIII HaeIII
cgtggaggacgacgtgtgcccaggagcccgcagcaggaggagtgccatggcgaggaggagccgaggagca base pairs
gcacctctgctgcacacgcggtcctcgggtgctcctcctcaaccggtaccgctcctcctcgggtcctcgt 1051 to 1125

HaeIII MspI HpaII 9eF
ggccatgcaggaggaggtggttgcggtcgccgggggtcaccgggaagaaaggggtggggaagaaagaaagccgc base pairs
ccggtacgtcctcctccaccaacgccagcgccgcccagtgggcctctctcctccacccctctccttgcggcg 1126 to 1200
ScrFI
SmaI

HpaII HhaI HaeIII
cgccgcgctggcgaccacgagccggagttcatcgccagccctgttgcgcggacgaggcgcgagcaactggcc base pairs
cgggcgcgaccgctggtgctcgccctcaagtgcggtcgggacaacggcgctgctccgcgctcgttgaccgg 1201 to 1275
MspI

HhaI HaeIII
caaagcgctacggccgcagcacttgcgcgaaaagaagtacattatcttcccagctctggttttgatttgacca base pairs
gtttcgcatgcccgcgtcgtgaacggcctttctcctcatgtaataaaagagggtcgagacaaaactaaactggt 1276 to 1350
9eR

HpaII
gattttactccatgtctgttagtacttgcgagctgagcaatctgctatttgcgtgatttattgtgcgtgcagacc base pairs
ctaaaaatgaggtacagacaatcatgaacgctcgactcgttagacgataaaacgactaaataacacgcacgtcctgg 1351 to 1425
MspI

SacI HaeIII HpaII HhaI
ggatgaggaggaagagctcaaggccagatgtcactaccggagcgctaaaggtggacaacgctcgtctactgcctcgg base pairs
cctactcctccttctcgagttccgggtctacagtgtatggcctcgagattccacctgttgcagcagatgacggagcc 1426 to 1500
MspI

EcoO109I
ggatgacgtctatgtcaaggtccttgttcatcgcttctgttgccttgcctcatttatgatgtgcatatgtgt base pairs
cctactgcagatacagttccaggaaagtagcgaaagacaacgaagacgagagtaataactacacgtatacaca 1501 to 1575
AvaII

MseI HinfI HpaII
ttgttaagggaagcaagaattgcttgattttgttgcgactcgcatctccgtgacgagttctgcgtatggtcacc base pairs
aacaattccttcttcttaacgaactaaaaacaacggctgagcgtaaggcactgctcaagacgcataccagtggt 1576 to 1650
MspI

ScrFI
TaqI BstNI Sau3AI
ggtagctggcactgatacacaacgtggtatgctggaagctggttagtatatttgcacgaccaggaggtccaga base pairs
ccatgcaccgtgactatgtgttgaccatacgaccttcagaccatcatataaaacgtagctgggtcctccaggtct 1651 to 1725
EcoRII AvaII

ClaI 16iF HinfI
tcgatatgtgcggtatagtgcttatttgattgcacccctgttcggagattcargcatgagtcggtgttagatgac base pairs
agctatacacgcatatcacgaataaactaacgtgggacaagcctcctaagtagctactaccgcacaaatctactg 1726 to 1800
TaqI

ScrFI BstNI
PvuII EcoRII PvuII HaeIII HpaII HhaI HinfI
gcctcccagacagctgcctgccaggcagctgattctggcccaggcgctccggaatggtgaagttgcgctggcaaga base pairs
cggagggtctgctcgacggacggtccgctcgactaagaccgggtccgagggccttaccacttcaacgcgacgcttct 1801 to 1875
BstNI HinfI EcoRII MspI
ScrFI

ScrFI
HaeIII EcoRII
ttctcaggccacctaccaaataatgccctggagcatattgcagcttcttttttgccttcttctctatattt base pairs
aagagtcgggtggatggtttatacgggacctcgtataacgtacgaagaaaaaacaagagaaaggaagatataaa 1876 to 1950
BstNI

atctcattgttagtgaagtttcacattgcagctgtcatggaatatttactttcaaatcaacaggagagatgctagc base pairs
tagagtaacaatcacttcaaagtgtaacgtgcacagtaccttataaatgaaagttagttgctcctctacagatcg 1951 to 2025

9/39

FIG. 5

EcoRV

Continued

attgaggtgtgtgataattattacataactagaagatatcgtgcatgttgccattgggattgcaagaatgtggaa base pairs
taactccacacactatttaaatgtatgatcttctatagcacgtacaacggttaaccctaacgcttcttacacctt 2026 to 2100

MseI MseI

atgatgttggttagcttgtattaaagaggttaacaggttagtgggatgacatgactattagttagagatgatgtggat base pairs
tactacaacacatcgaaacataattctccaattgtgcaatcacccactgtactgataatcaatctctactacacctt 2101 to 2175

agtaagtgggatatgatgtagatgacttgtgtgttgagacagaactataacatggaggtggaaatgggagcagca base pairs
tcattcacccctatactacatctactgaaacacacaactctgtcttgatattgtacctcaacctttaccctcgctgt 2176 to 2250

MseI

tggtcaaacataccctaaatgcctgtctctacacaatgtgggtgattgggtgtatagctgtgtttaaagctggat base pairs
accagtttgtatgggatttaccggacagagatgtgttacaccactaaccacatatcagaccacaattttcgacctt 2251 to 2325

HinfI

MseI

XbaI

ScrFI

EcoRII

actttgattctgttgaagattgtcacacccgaatttaaggacaaatctagatacatctcatatgtgcaccaggat base pairs
tgaaactaagacaacttctaacaggtgtgggcttaaatcctgttttagatctatgtagagtatacacgtggctcta 2326 to 2400
BstNI

agtgtatagataccaatgtcataatctttattacacgacgataatgtcttcaaaaatctgggtgttacaagatg base pairs
tcacatatctatggttacagtattagaaataatgtgtgtctattacagaatgttttatagaccacaatgttctac 2401 to 2475

MseI

MseI

cacctttcaacatgttttaagtctgcaaaactgttttaattaaacagaatgcagtggtttgaacaaaaaatgtctgc base pairs
gtggaaagttgtacaaattacgacgtttgacaaaatttaattgtcttacgtcacaaaactgtttttttacgacg 2476 to 2550

MseI

15iF

Sau3AI

HinfI MseI

tttatcctgcacatctgttttgcagtggtcagcaattgcttacattccattatgatctctgagattctttaaattt base pairs
aaataggacgtagaacaaaacgtacacagtcgttaacgaatgtaaggtaatactagagactctaagaaatttaa 2551 to 2625

ctagcatgatgaaagtatttactaattcaactgaacacaaacattgtttgaatgaacaaggcaacacggatgctt base pairs
gatcgactactttcataaattgattaagttgacttgtgtttgtaacaaacttacttgttccgttgcctacgaa 2626 to 2700

MseI

ggaataatgggtgtgtataatatcacttagtggttttgcctcacaccacatctttcatgggttctttaataata base pairs
ccttattaccaacacatatattatagtgaaatcacaaaacgagagtggtgtgtgaaagtacccaagaaattattat 2701 to 2775

MseI

HaeIII

gttactgactttaagtttcttattccttttgtctatcttagctggagaaaacgaggcagattacattggccgc base pairs
caatgactgaaattcaaagaataaggaaaaacagatagaatcgacctcttttgcctcgtctaagttaaccggcg 2776 to 2850

14eF

attactgaattttttgaggggactgaccagtgtcactattttacttgccttgggttcttccgagcagaggacacg base pairs
taatgacttaaaaaactccccgactggtcacagtataaaatgaacgggaacaaagaggtcgtctcctgtgc 2851 to 2925

15eR

gtgtgtatttagtattttgtcattctatgcatgtgtggatttttctggaatgtggaaaacatacagcactctctc base pairs
cacacataaatcataaaacagtaagatacgtacacacctaataaagaccttacacctttgtatgtcgtgagagag 2926 to 3000

MseI HaeIII

HaeIII

tacaccacacacacttctagtatatgtgtacacgttaattgggccaacactagacacatggcccaacatccccct base pairs
atgtgggtgtgtgtgaagatcatatacacatgtgcaattaccgggtgtgtgatctgtgtaccgggtgtgaggggga 3001 to 3075

EcoRV

caagatgggcatagatatcaatcatccccatcttgctacataacacatcacactcttttactcctataccctta base pairs

10/39

FIG. 5

Continued

gttctaccgcgtatctatagtttagtaggggtagaacgatgtattgtgtagtgtgagaaaaatgaggatatgggaat 3076 to 3150

HinfI

gtcaagcaatctgctattttgaccttttgagtttacatgattcaactctaaagtaccattatctaacttctctttg base pairs
cagttcggttagacgataaaactggaaaactcaaagtactaagttgagatttcatggtaatatagattgaagagaaac 3151 to 3225

ClaI

HinfI

atgaagaatcgatcaatttccacatgttttgttctatcatgttgaaactggattgttagctatattcatggctgac base pairs
tacttcttagctagttaaaggtgtacaaaacaagatagtacaacttgacctaaacatcgatataagtagcactg 3226 to 3300

TaqI

Sau3AI

HinfI

MseI

18eF

ttattatcacaccataacttcaggaggtcttttcttaatacattcaactctgataagagaccctttatcagatagc base pairs
aataatagtggtgattgaagtccttcagaaaagaattatgtaagttgagactattctctgggaaataggtatcg 3301 to 3375

HaeIII

atctcacatarcgcaaggccatagctcggtattctgcttcggcggtggaacgggataccacagattgtttcttg base pairs
tagagtgtatagcgttcccggtatcgagccataagacgaagccgccacttgccttatggtgtcttaacaaagaac 3376 to 3450cttctccatgatactaaatttctccaacaaacacacaatatctcgaagttgaccttctatcatcaaggcaacta base pairs
gaagaggtactatgatttaaaggaggtgtttgtgtgttataggacttcaactggaagatagtagttccgttgat 3451 to 3525

MseI

SbfI

EcoRII

ccccagtcgtcatcagagtaaccttccaccttttagatgacctgacctttaaagattattccctttccaggacaa base pairs
ggggtcagacgtagtcctcattggaaggtggaatctactgggtactggaatttctaataagggaaggtcctgtt 3526 to 3600

BstNI

TaqI

gtcttcaagtatcgagtatagcatcactgcatcaagatgtccacttctggggtcatgcatatatcgactcacc base pairs
cagaagttcatagcgtcatatgcatgtgacgtagttctacaggtgaagacccagtagtatatagctcagtggt 3601 to 3675

HinfI

EcoRV

acactgactgcatatgtgatatcaggtcttgtatggcacaagtagatgagccgtccaacaagtctttgatacctt base pairs
tgtgactgacgtatacactatagtcagaacataccgtgttcatctactcggcaggtgtttcagaaactatggaa 3676 to 3750

Sau3AI

HinfI

HinfI

TaqI

HaeIII

tctttattcacaggatcaccagattcagcacataatttatgattcaagtcgataggtgttgctacaggccgacac base pairs
agaaataagtgtccttagtggtctaaagtcgtgtatataactaagttcagctatccacaacgatgtccggctgtg 3751 to 3825

Sau3AI

Sau3AI

cccaacatacctgtttcatcaagtagatctaaacatatttctttgggagagaactattccttttgagatcga base pairs
gggttgtaggacaaaagtagttcatctagattttgtataaaaggaaacctctcttgataaggaaaacctctagct 3826 to 3900

BglII

TaqI

Sau3AI

HinfI

gcaatctcaataccaagaaagattttgagatgaccaagatctttaacctcaaattccttacttagattcttcttt base pairs
cgtagagttatggttctttcataaactctactggttctagaaattggagtttaagggaatgaatctaagaagaaa 3901 to 3975

BglII MseI

Sau3AI

agacatgcaatctcaagatcggcatcacctgttaataataatcatccacatacacagtagaattgcaattcgt base pairs
tctgtacgttagagttctagccgtagtgacattattattatagtaggtgtatgtgtcgatcttaacgttaagca 3976 to 4050

Sau3AI

cgctccaaagtgttgataaaaaacagtggtgatctccgttgcatgttttatatcccatgctacatattgcacgtcta base pairs
gcaggtttcacaaactattttttgtcacactagaggcaacgtaacaaatatagggtacgatgtataacgtgcagat 4051 to 4125

TaqI

aatctgtcaaaccatgctcttggggactgcttgagaccatacaatgatttttcaatcgacaaactttcccaatt base pairs

11/39
FIG. 5
Continued

ttagacagtttggtacgagaacccctgacgaactctggtatgttactaaaaaagtttagctgtttgaaaggggttaa 4126 to 4200

ScrFI
EcoRII Sau3AI

gtctcaggctttgacaatccaggaggatctccatataagacctctcttgcaaatcaccatgtaagaaagcattc base pairs
cagagtcggaactgttaggtcctccctagaggtatatctggaggagaacgttttagtggtacattctttcgtaag 4201 to 4275
BstNI

MseI HaeIII Sau3AI
ttaacatctagttgatacaaggccatccaaaatttgagcacaagagatcaatgtccttacagtactcattttt base pairs
aattgtagatcaactatgttcccggtaggttttaaacgtcgtgttcttagttacaggaatgtcatgagtaaaaa 4276 to 4350

gccactgggtgcaaatgtctcatcataatcaattccatatgtttgactataacctcttgcaaccaatcttgcttta base pairs
cgggtgaccacgtttacagagtagtattagtttaagggtatatacaaaactgatatgggagaacgttggtagaagcgaat 4351 to 4425

tatcgttctacccttctctctgggttttgcctcacagtgaatacccatttacaactaactgccttctttctttta base pairs
atagcaagatgggaaggaagacccaaaacgaagtgtcattatgggttaaatgttgattgacggaagaaaggaat 4426 to 4500
18iR

XbaI MseI
ggtagtttctcaaatcccaagtttgattttttcttagagctttaagctcctccaacattgcctcacgccagtta base pairs
ccatcaaagagtttaagggttcaaaactaaaaaagatctcgaaattcgaggaggttgaacggagtgcggtcaat 4501 to 4575

gaattacattgtgcttcttccaatctcttggaattgctacggaatgcaatgatgcaacaaatgctctatatgat base pairs
cttaatgtaacacgaagaaagggttagagaaccttaacgatgccttacgttactacgttgtttacagatatata 4576 to 4650

HinfI
ggtgacaaagacgcataatgagacataattgctaatgtcatgttcataatccataaccttgttgggggagctccagct base pairs
ccactgtttctgcgtatactctgtattaacgattacagtacaagtatagggtatggaacaacccccctgaggtcga 4651 to 4725

HhaI
ttagcacgcgctccttttctgtattgcaatgggcaaatcataagtgtcataatcttctcagtttctccatgagacgtc base pairs
aatcgtgcgcgaggaaaagcataacggttaccggttttagtattcacagtattagaagtcaagaggtactctgcag 4726 to 4800

aaaggtagatttatagcctctaattgtgtttggagagaactgctcagtagctgaattggtttcaggagacc base pairs
tttccatgtaaatatcggagattacacaaacctctcttgacgagtcataaactacgacttaaccaaagtcctcgg 4801 to 4875

tgagggtgcacatgggactttcttctgtatatacttcgaccttatatcgtaagtcgtctccacaagatttatta base pairs
actccaacgtgtaccttgaaagaagacatatgaagcgggaatatagcattcagcagaggtgttctaataaac 4876 to 4950

ttctcgtgactaggatgtgtctccaattcacttggcattacttgcattcttttgagaagcaccaatcaccacttcc base pairs
aagagcactgactctacacagaggttaagtgaaccgtaatgaacgtagaaaactcttctggttagtggtgaagg 4951 to 5025

HinfI TaqI
attttatttgggtgtgttccattgaaatcaaccattctgttctccccctctcgactagcttcatctgtgctagta base pairs
taaaataaaccacaacaagggaacatttagttggttaagacaagagggggagagctgatcgaagtagacacgatcat 5026 to 5100
16iR

HinfI Sau3AI
gagacagaatcaagaaaaaatttagatctgtcttctcaccatagaaaggcacagtctctctaaatgtaacatcc base pairs
ctctgtcttagttcttttttaaatctagacagaagagtggtatcttccgtgtcagagagatttacattgtagg 5101 to 5175
BglII

HinfI PstI
atgcttacaacaaacgtcgttccactaggactccaacacttgtatcccttttgccctgcaggatatacaaaaaa base pairs
tacgaatgtttgttgcagcaagtgatcctgaggttgtgaacatagggaaaacgggacgtctctataggttgtttt 5176 to 5250
EcoRV

BamHI Sau3AI
atgcacttcacagcagaggtccaaactccccacctgaggtctatgatctctgacaaaacatgtacatccaaaa base pairs
tacgtgaagtgtcgtctccttaggttgaaggggtggactccagatactagagactgtttgtacatgtaggtttt 5251 to 5325

Sau3AI

12/39

FIG. 5

Continued

HinfI

HinfI

atttttaggtggaaccacaaaacttattctcaccgagaagaatctcacatggagtccttcattgcaagtatttttgaa base pairs
taaaatccaccttgggtgttgaataagagtggctcttcttagagtgtacctcagaagtaacgttcataaaaaactt 5326 to 5400

MseI

ggagtgcgattaataagatatgtggcagtcataacagcttcactccataggaacttcggaacattcattgtaaac base pairs
cctcagcgaattattctatacaccgtcagttatgtcgaagtgggtatccttgaagccttgaagtaacatttg 5401 to 5475

14iR

HinfI

atcagcgaacgagcaacttccaaaatgtgacgattcttcttccagccacaccattttgtggaggtgtatcagga base pairs
tagtcgctgtcgttgaaggttttacactgctaagaaggaaagtcgggtgtggttaaaacacctccacatagtcct 5476 to 5550

MseI

caggatgtctgatgtaataaccattcttctgacagaaatgcattaaatcccttgtttacataactcggttcattg base pairs
gtctacagactacattatattggtaaagaactgtctttacgtaatttagggaaacaaatgtatgagccaaggtaac 5551 to 5625

11iF

HinfI

tctggtcttaggattttgacttgagttgaattgattctcaactagtgccacaaaaattttgaaaacacttcaat base pairs
agaccagaatcctaaaaactgaactcataacttaactaagagtgtatcacgtgtttttaaaacttttgaagttta 5626 to 5700

12iF

TaqI

acttcattctttatgcttcacacatagacccaagtcattccgagaaaaacaatcgataaagtaacaaagtacttc base pairs
tgaagtagaaatcgaagtagtgatctcgggttcagtaaggctctttttgttagctatttcattgtttcatgaag 5701 to 5775

ClaI

MseI

atccattaatagaagtcacaggacatgtccaaacatcagaatgaactagcacaaaaggagatatactcctgata base pairs
tagggttaattatcttcagtgctcgttacaggtttgtagtcttacttgatcgtgttttctctatatagaggactat 5776 to 5850

TaqI

cctcgactaataataagatgtccttgtgtgttttgcaaactcacaggcatcacacaatagcttgcttttatccacc base pairs
ggagctgattatattctacaggaacacacaaaacgtttgagtgctcgtagtggttatcgaaacgaaaataggtgg 5851 to 5925

HindIII

ccactcattacatcaggaagcctttgcatatcttatcaaaagaaagatgccttaactcacaatgcaagagcatc base pairs
ggtagtaaatgtagtccttttcgaaacgtatagaatagtttcttctacgggattagatgttacgttctcgtag 5926 to 6000

Sau3AI

actgcaacctccttcttcttccattcttgttgcagcatagtgcatattgtaccattagtcctctcatgatccata base pairs
tgacgttggaggaagagaaggtaagaacaacggtcgtatcacgtataacatggtaacaggggagtagtaggtat 6001 to 6075

ScrFI

EcoRII

MseI

taccacaatccattacgccttgtagctgtcccaagtctcttccctgtttccctctcctgaattaaacaattatct base pairs
atggtgttaggtaaatgcggaccatcgacaggggtcagagaaggacaaaggagaggacttaatttgttaataga 6076 to 6150

BstNI

TaqI

Sau3AI

EcoRV

cgatcaagaataatcagacaatccaattgatcaaccaaggcacttagtgatatcaagttgacaggaaagggtggc base pairs
gctagttcttattatgtgttaggttaactagttgggtccgtgaatcactatagttcaactgtcctttccaaccg 6151 to 6225

Sau3AI

MseI

acatacaaaactgatgacaacttaatagatggagtgcattgcactgtgccaacacccttgatgggttgggtgta base pairs
tgatgttttgactactgttgaattatctacctcacgtaacgtgacacgggttggggaactacccaacaccacat 6226 to 6300

EcoRV

ccatcagcagttgtataatttctttacgtgtgggggatatttatatgatgtaaaattcactggacgtgcct base pairs
ggtagtcgtcaaacatattaaagaaatgcacacccccctatagaatatatactacatttaagtgaactgcacgga 6301 to 6375

13/39

FIG. 5

Continued

HinfI MseI
 gtgacatgctttgatgctcctgagtcctaaaatccattttaactgtgtgacctgtgtgggtacaaaagcatgagca base pairs
 cactgtacgaaactacgaggactcagatttttaggtaaaattgacacactggacacacccatgttttcgtactcgt 6376 to 6450

HinfI Sau3AI
 taattaccttcatcagtgtagggcaagtggacaaaaatccctgtgtgagactcctgatctttatctccagagatt base pairs
 attaatggaaagtagtcacatccgcttcacctgttttaggggacacactctgaggactagaaatagaggctctctaa 6451 to 6525

tgatttttcttccctcaactttgtttcatcttcgtgttccataaatgtttcaagttcttcttgtgtagtgtgtgca base pairs
 actaaaaagaaggagttgaaacaaagtagaagcacaaggtatttacaagttcaagaagaacacatcaacgacgt 6526 to 6600

MseI
 ttgcctcttgcccaactcctgcctccacgacctctgccgctctaggagccctcttctctccacagattaact base pairs
 aagcgggaacgggttgaggacggaggtgctggagacggcgagatcctcggggagaaggagagggtgctaattga 6601 to 6675

MseI
 ttggaaggcttagaacaattacgtgcaatatgtccaacattaccacaattgtaacattctctagtatctttggtt base pairs
 aaccttccgaatcttgttaatgcacgttatacaggttgtaatggtgttaacattgtaagagatcatagaaaccaa 6676 to 6750

ScrFI
 HinfI HinfI EcoRII
 ctcatagctgaaaaacacaggatgaggcggcgctttgagaactttctctcatcacttttgagtccttgactcctcctgg base pairs
 gagtatcgacttttgtgtcctactccgcccgaactcttgaaagagagtagtgaaactcagaactgaggaggacc 6751 to 6825
 BstNI

TaqI
 gatatggcagctatggcttctttaggctaggaagagtgaggtgatgaaacatggaggcacgtcttccctcgaac base pairs
 ctataccgtcgataccgaagaacatccgatccttctcacctaactttgtacttccggtcagaaagggagcttg 6826 to 6900
 13iR

tctgagtttagcccccttagcaattgaagtacacgtcttttttccacccatttcttccgccaagcaacacactct base pairs
 agactcaaatcgggggaatcggttaacttcatgtgcagaaaaaagggtgggttaaagaagcgggttcgttgtgtgaga 6901 to 6975

Sau3AI Sau3AI
 gagtgtggtagctcaataggatcataatgatcaacatcagcccataaacattgtaactcctgaacgtactccgcc base pairs
 ctcacaccatcgagttatcctagttactagttgttagtcgggtatttgaacattgaggacttgcatgaggcgg 6976 to 7050

Sau3AI Sau3AI 13iF
 acagatcgctccccctgtttgatattatggaggcagctctcagctctccaccatcaacataacatttccagctccc base pairs
 tgtctagcgagggggacaaactataatacctccgctagaaagtcagaggtggttagttgtattgtaaaagtcgaggg 7051 to 7125

PstI HinfI
 gagtacattttctcaagtgtctttccacatttctgcagcacttatgattgtatcaacagtgctagcaattgctgga base pairs
 ctcatgtaaaagaagttcacgaaagggtgaaagacgtcgtgaataactaactagttgtcacgatcgttaacgacct 7126 to 7200

MseI
 atcatagaactcaacatccacgtgccactaaagagtttatagcatcccagttcttccattcatcacttaactta base pairs
 tagtatccttgagttgtaggtgcgacggtgattttctcaaatatcgtaggggtcagaaaggtaagtagtgtaattgaat 7201 to 7275

HinfI XhoI XbaI
 MseI
 tccttggggtcaacgacatctccttttaacatagccctcgagtccttttgccttcaataatcgcaatgctcttcta base pairs
 aggaaccgcgagttgctgtagaggaaattgtatcgggagctcagagaaacggaagttattagcgttaccgagaagat 7276 to 7350
 TaqI S4iR

MseI
 gaccatgccaaataatttttccccccttctaacttaattctcatttggcattaggtctatcttctgaactgggtct base pairs
 cgggtacgggtttattaaaaagtggggaagattgaattagagtaaacctgaatccagatagaagacttgaccaaga 7351 to 7425
 12iR

MseI TaqI
 atatgagcaacattgtcttttaattgatgagggcctcatccctttttgtgtgacagtaattcgaccaatttacca base pairs
 tatactcgttgaacagaaattaactactacctcggagtagggaaaaaacgactgtcattaagctggttaaatgggt 7426 to 7500

PvuII

15/39

FIG. 5

Continued

BamHI

PvuII

ataatgtttctgactaccattgttttggctacttgccttagatggatccaaaagccaaggctcagctgatag base pairs
tattacaaagactgatggtaacaaaacacggatgaacggaatctacctagggtttcgggtccgagtcgactatc 8401 to 8475
Sau3AI

agagtgcgacctatactatgacatgtcttactctgttgcataattctacatttgctaataatctcgtctggaatt base pairs
tctcaacgctggatgatgactgtacagaatgagacaaagcgtataagatgtaaacgattatagagcagaccattaa 8476 to 8550

MseI

ccttctgcacatcttttttgggtgactagctgaatgcaggttagccttgcgaagaggttaatacatgagttgtt base pairs
ggaagacgtagtagaaaaaaccaactgatcgacttacgtcaatcgaaacgggttttccaatttatgtactcaacaa 8551 to 8625

TaqI

MseI

MseI

cctgcactcgaagggatgtcaataatgtccacaaactctgaaaatgtatttttagatacttaacttggtaagt base pairs
ggacgtgagcttttccctacagttattacaggtgtttgagactttacataaaaatctatgaattgaacaattca 8626 to 8700

cagtaaacctgtcagatacttgggttttgggtacgattaccatccttatgtgagtaaacctcgtcaagggatgt base pairs
gtcattttggacagctctatgaacccaaaacccatgctaatggtagaataacactcattttgagcagttccctaca 8701 to 8775

TaqI

Seq2FN

caatgacgtgttgattgtgtattagatattctgtttgttcgaagcraatctacacaaaacagcttatgtaattga base pairs
gttactgcacaaactaacacataatctataagacaaacagctttcgattagatgtgttcgtcgaatacattacat 8776 to 8850

HindIII

HaeIII

aaacctcaaaacaaacttgccctcttcataagcttaggtttataggattagcgttttagtgcagtgaaggcctatttg base pairs
tttggagtttgtttgaacggagaagtattcgaatccaaatatacctaatacgaatacagtcacattccggaataac 8851 to 8925

HaeIII

BstNI

SacI

ScrFI

EcoRII

TaqI

ScrFI

EcoRII

cttcacggcctccctgccgagctcctgggttagacagccatcctggccgtaggtgcccgaatcgaacacctggga base pairs
gaagtgcgggagggacggctcgaggaccgatctgtcggttaggaccggcatccacgggttttagcttgtggaccct 8926 to 9000

EcoRII

BstNI

ScrFI

HaeIII

BstNI

ScrFI

EcoRII

gccacgtttgcactagcaggttttctcgtgggtgcaaaccaaacacgcctatagtgttcaagtataactgaattggt base pairs
cggtgcaaacgtgatcgctccaaaaggaccacgtttgggtttgtgcgggatcacaaagttcatattgacctaaacca 9001 to 9075

BstNI

MseI

gctcacctttgtcctaagcttaagtttttgggttttcatcgggtgcagtgcactccatactcaatagtcfaatga base pairs
cgagtggaaacagattatcgaattcaaaaacaaaagtagccacgtacgttgaggtatgagttatcagttatact 9076 to 9150

BstNI

XhoI

HinfI

tatagtgttcaagcatagaactctcgagtttgaatcctggcaggggcaatcaataaaaataattgcagcttacc base pairs
atatcacaagttcgtatcttgagagctcaaaacttaggaccgtccccgttagttttattttattaacgtcgaatggg 9151 to 9225

TaqI

EcoRII

ScrFI

SfiF

ctattttctacgtttgagcacatgagggagaggttgaattataagtgtgttctccatctttcttaacagatgaa base pairs
gataaagatgcaaacctcgtgtactccctctcacaaacttaatttcacacaagaggtagaagagattgtctactt 9226 to 9300

HinfI

MseI

MseI

ctgggtttgtgcagtgaactcaatatgatatttgagtcaaatgtttactttaaaatcatagttgatgcaattta base pairs
gaccaaacacgtacattgagttatactataaactcagttttacaaatgaaatttttagtatcaactacgttaaaatta 9301 to 9375

16/39

FIG. 5

Continued

aacatatttttttggctcgtgtagggagtgtaggtataactgaattgcacacatttccttagcttaggttt base pairs
ttgtataaaaaaacagagcacactccctcatgcatattgacttaacgtgtgtaaaggaatatcgaatccaaa 9376 to 9450

Sau3AI

ttgactgcaactgttggtgcatgtagctcaataactaaagtgtgacagctctacagtgaataagtttgaca base pairs
aactgacgttgacaaccacgtacatcgagttattgatttcaactagacctgtcagatgtcacttattcaaactgt 9451 to 9525

cttgtaaaatgtgcatgtatttttacaacgctggcacttttttccaatagaaaatgggcagtcaggcagtgat base pairs
gaacattttacacgtacataaaaatgtttgagccgtgaaaaaaggattatcttttaccgctcagtcgctcacta 9526 to 9600

Sau3AI

accgcttcgggtattttctctgtagtggtgtagctgtagctatgtagtgcacacgaggacagcaaccctt base pairs
tggcgaagcccataaagaagactactacacctagacctctgcagtagatcatcaggttgcctgtcgttgggaa 9601 to 9675

Sau3AI

cttgatctgtattctggctgtggggcatgtctactggtctttgcttgggtgcagctctttctggcttgaaactt base pairs
gaactagacataagaccgacaccccgtagacatgaccagaaacgaaccacgtcgagaaagaccgaactttgaa 9676 to 9750

Sau3AI

gaaactgtaattcttcaactagtcattctgttgtagaataatgttcacgatctcagaacttattctattgttctg base pairs
ctttgacattagaagattgatcagtagacaacctatcttatacaagtgtcagtagcttgaataagataaacaagac 9751 to 9825

MseI

gcttgacgagtggtggtgtgatttcaacagttttgctgccaagtttaaaatataatcatccacagactgagg base pairs
cgacgtcgctaccgcacaactaaagtgttgcacaaacgcaggtttcaaatcttattagtaggtgtctgactcc 9826 to 9900

HinfI

tatggatagtaaaacttcatcttggattccatctgttctgtcagctactcttacaagtgcttggtatttttgatg base pairs
atacctatcatttgaaagtagaacctaaggtagacaagacagtcgatgagaatgtttcacagacctaaaaacctac 9901 to 9975

MseI

taggtgcgaaatgagaaagccgatgagtttcttgcctccttaaggaatgggcagttctatgcaaaaaatattgc base pairs
atccacgctttactctttcggtactcaaagaacgggaggaattccttaccggtcaagatacgttttttatcacg 9976 to 10050

HinfI

Sau3AI

caagatgtggattcaaatttagcaagctcagaggatcaagcggatgaagacagccctcttgacaaggacgaattt base pairs
gttctacacctaagtttaaatcggtcaggtctcttagttcgctacttctgtcgaggagaactgttctctgcttaaa 10051 to 10125

Seq2RN

HindIII

MseI

gttgtagagaagcttgcgggatattgttggtggcagtgacagggaaaaatggcatctattttaaggtacttcag base pairs
caacatctcttcgaacagccctatacaataccaccgtcactgtcccttttaccgtagataaaattccatgaagtc 10126 to 10200

HinfI

MseI

tgtcatttgttcatttctacttgattccaacaaaaaatcaattacttaagcctgtcaaacgatggatatttctg base pairs
acagtaaacagtaagatgaactaaggtgttttttagttaatgaattcggacagtttgctacctataaaagac 10201 to 10275

PstI

HaeIII

tatatttgcgtaacgctagatttctgcaggtccagtgaggagatacggccctgaggaggatacatgggaacc base pairs
atataaaacgacatttgcgatctaaagacgtccaggtcaccttctctatgcccggactcctcctatgtacccttgg 10276 to 10350

AvaII

gattgataacttgaggttagtgatggtatatcgctctgttgttgccttgtatacctatttgcattctaactcttg base pairs
ctaactattgaactccaatcacataccatatagcagacgaacacggaacatattggataaaacgtagatttaggaac 10351 to 10425

Seq1RN

HinfI

FIG. 5

Continued

S3eR

ECORV

HaeIII

MseI

3F

3R

cgtagcttgaagcccaagatgtgtctcatggaaaaatgtggtggacatactcaaaatttgcggatggtacctaggaa base pairs
qcatgaaacttcgggttcatacaagagtacctttacaccacctgtatgagtttaaacgcctaccaatggatcctt 10726 to 10800

Khái

1F

Race2A

Race2B

ScrFI

EcoRII

HinfI

BstNI

Race 1A

4F

Race1B

RaceRT

cacaaccactactatgaaatcatgtggaatgtgtaaaatacgcgtgaccaactgaatttgttgacgaatgtatg base pairs
gtgtttggtgatgatactttagtacaccttacacattttatgcgactggttgacttaacaacgcgttacatac 11101 to 11175

gttgcataatgacgagacacaaaaaaccatccctgaaaaagccttgcttcttggcgatgcaatttcagatttacca base pairs
caacgtatactgctctgtgttttggtagggactttttcggaacgaagaaccgctacgttaaagtctaaatggt 11176 to 11250

M96T

Pst I

EcoRI

Hha I

TagI

HinfI

HaeIII

MspI

Pat I

Avail

18/39

FIG. 5

AvaII

Continued Sau3AI

MseI

cttcggtgaagggggtggtccagatgaaggcaagctcttgatcaccagcctttacggttaacaacgatgat base pairs
 ggaagccacttccccgaccaggtctacttccgttcgagaacctagtggctcggaatgccaattgtgtacta 11551 to 11625

HinfI

tatgagcgggttcaacagattcctgtcaagaaggttggtggcttgctgcatttgccttccctttgtgtttt base pairs
 atactcgcccaagtgtgttaaggacagttcttccaaccaccgaaccagcgtaaacacggaaggaaacaacaaaaa 11626 to 11700

5F

tccccctctgaacaatcatctctcttctctatgacagggagccaacttccgcgacctaaagggcgtgaggggttg base pairs
 aggggaagactttgttagtagagagaaggatactgtccctcggttgaggcgctggatttccgcactcccaac 11701 to 11775

BamHI

TaqI

gagcaacaatattgttgagtggtccagaaatcgagcgtgtgaaactttcatctgggaaaccactggatgtg base pairs
 ctcggttgttataacaactcaccctaggtcttttagctcgacactttgaaagtagaccctttggtgaccatacac 11776 to 11850

Sau3AI

tgctatttccgtgctgtgttttctataactgtgcaacatttactttcccatattcaaaactcataactgacgaga base pairs
 acgataaaggcagcacaacaaaggatattgacacgttgtaaatgaaaggtataagtttgagtattgactgctct 11851 to 11925

HinfI

tgctgcaactactgtgaagattcatggctaaccatgacaacattttgcacacatctttgttatctaggttctga base pairs
 acgacgttgatgacattctaaagtaccgattgggtactgtgtgtaaaacgtgtgtgaaacaatagatccaaggact 11926 to 12000

ctatgcaatgtcattcatcaaggcacaatcactcaagtaagtttcaaacattttttgttttttgggggaaaa base pairs
 gatacgttacagtaagtagttcccggttttagtgagttcattcaagttttgtaaaaaaacaaaaacccctttt 12001 to 12075

HaeIII

HhaI

gtaggttattgtttacttgtgtttacatgatgtgtgcaggccgtttggggcctgtgtgtgggacaagacagttc base pairs
 catccaataacaaatgaacacgaatgtatactacaacgtccggcacaaccgcggacaccacctgttctgtcaag 12076 to 12150

ScrFI

EcoRII

HaeIII

ctacagttgtaaccagagcagagcctcacaccaggtcagcttcagaaaggccactccttttcgccaatccctgc base pairs
 gatgtcaacattggtctcgtctcggagtggtgtgtccagtcgaagtctttccggtgaggaaaagcggtagggacg 12151 to 12225

BstNI

Sau3AI

atctgtatttactattagcgtgtgttcccatatgatcattaccgaacatgtgttccacacaggttataattcatc base pairs
 tagacataaatgataatcgacacaagggtatactagtaatggctgtacaacaggtgtgtccaatattaagtag 12226 to 12300

ScrFI

HinfI

EcoO109I

HpaII

cgactcaagcaaggggtcctcactatccgggagaacgcaaggttacagggcttccccgattattaccgattgtttg base pairs
 gctgagttcgttcccaggagtgatagccctcttgcgttccaatgtcccgaaggggctaataatgggttaacaaac 12301 to 12375

AvaII

MspI

HaeIII

Sau3AI

gcccgatcaaggagaagtaagttcctgttttcaagttgcctgtaccagatctagtactattgaaagttttcagc base pairs
 cgggctagttcctcttcatcaggacaaaagttcaacggacatggctctagatcagtgataactttcaaaagtcg 12376 to 12450

Sau3AI

BglII

agcaagccattcatcagttagttacagctcttgaaagccttacctctgaacatgtgtgtttctctgtatgggtgat base pairs
 tcggttcggtaagtagtcaatcaatgtcgagaactttcggaatggagactgttacacacgaaagagactaccacta 12451 to 12525

MspI

HpaII

EcoRII

aggtagattcaagtcgggaacgcagtggtgtccctgttgcgggactgggtactgtctggggcaagcctac base pairs
 tccatgtaagttcagcccttgcgtcacgcaggggacaacgggcccgtgaccgatgacagaccccggttcggatg 12526 to 12600

ScrFI

SmaI

19/39

FIG. 5

Continued

ScrFI

HinfI

PvuII

ctgggtgaatctgaggggagtgacctctgtaccagctgcctccaagtttcacctctgttggaggacgcactgcg base pairs
 gaccacttagactccccctcactgggagacatggtcgacggaggttcaaagtggagacaacctcctgctgacgc 12601 to 12675
 BstNI

EcoO109I

PstI

Sau3AI

gggcaggcgagggcctcttcctgttggcaccctgcaggggaggttagttgagcagtaaaaggatgacagatctga base pairs
 cccgtccgctcccggagaaggacaaccgtggggacgtccccctccatcaactcgtcatcttcctactgcttagact 12676 to 12750
 HaeIII 1R BglII

TaqI

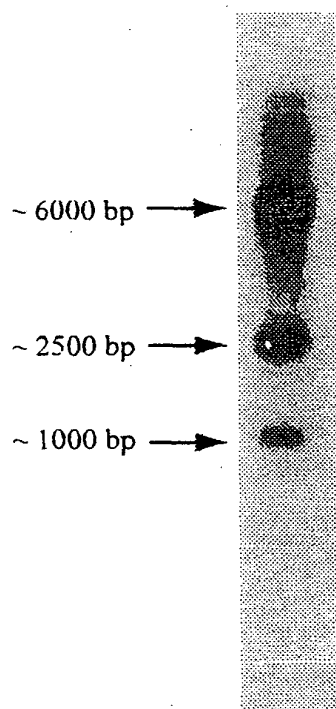
gctgagctgggcaacatccagcggcaggagcatttctggttcggttcgattcgggctcacga base pairs
 cgactcgaccggttgttaggtcgccgtcctcgtaaagaccaagccaagctaagccccgagtgtc 12751 to 12812
 HinfI

FIG. 6

PROCESS	WORLD WIDE WEB SITE
sequence format conversion	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/readseq.html
reverse complementation	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/revcomp.html
sequence translation	http://dot.imgen.bcm.tmc.edu:9331/seq-util/Options/sixframe.html
protein information	http://www.expasy.ch/tools
sequence alignments using Clustal W	http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html
sequence comparisons using BLAST 2.0	http://www.ncbi.nlm.nih.gov/gorf/bl2.html
sequence searches using BLAST 2.0	http://www.ncbi.nlm.nih.gov/blast/blast.cgi?Jform=0

21/39

FIG. 7



22/39

FIG. 8

zmet2 1 mapspapoptivsgkrkaakaeelhqpkkeeevvaasakrkaasgqtkpapkqakpgkkgkdaemkpvddvcaeeepdeeealMGeeeeeeQamQeEvavaAGSpq
 CMT1 1 -----MAnkqkkaEpsdlcfAGks-
 NLS
 zmet2 118 KKivgrinAaaagdhPePigspvaadEASnWPrYGrstaAKGDEEeELKARChYrSAKVDNvVYCLGDVVVAgeneadYIGRIteFEGdDqchYtCRWFFraEDTVIn
 CMT1 23 -----mSVvestiRwPhrydskktkIQApTtkPankG---GKGdEELIKAKCHfKALVdGVVLnLdDdVTVtQlpgkIKfAKVIELfEAdDgVpYICrfrWYrPpEDTLie
 NLS
 zmet2 235 slvasvdkhdkPRKRVtLseEKNDNVLDciISRVKivHVd-PndkPakaqfiesCDIYIDMSisvaxstkenisENGsgGSDtAGsISDdvdKetassmptrttatLIDYISCC
 CMT1 130 rfs-----HlvqPKRVtLsnDnONpLtiCiwSKVniAKVpLpKItshieqrVippcdYIDMktvPlnlt--aDd---GSDaSSLSGD-salnctenLhkdekLIDYISCC
 zmet1 1084 MET1 1096 LDIFAOC
 Motif I
 zmet2 351 GCSNtGICLALYGLKLETRMAVDNsfACQSLVNHPEtVRNEKADEFIALIKKwAVLCRY-----VqVDanLasseDQAdedsp---LokdEfvVVKIVGicYg
 CMT1 228 GANSTGICMGASISGVKLITKMSVDINKFACSLKINPEtVRNEKADEFIALIKKwAVLCRY-----VqVDanLasseDQAdedsp---LokdEfvVVKIVGicYg
 zmet1 1091 GGLSeglqQAGvsf-----TKWALIEeepAGaFinkHPEAVV---fVdncnVILKai-----MdkcQpTdcvSt-----se
 MET1 1103 GGLShGLKkAGvsd-----AKWALIEeepAGaFinkHPEAVV---fVdncnVILKai-----MdkcQpTdcvSt-----MdkgQpTdcvSt-----te
 Chromodomain
 zmet2 454 SD---REngIvKvQNGEGspEVPFIDnLadCpQIRetVQEGhKtKILPREDVDVVICZEPFCQIGTNYRNrDePLKDKNKQMVtFMDIVGLPKVIXENNVVILKVA
 CMT1 352 PqglqEhtLdImvWKIGinsvDTPWEPysqLncKtEKLKvYIDGfKSHLPpGTYtVCGPPCQIGIGYNYRNnEAPLEDQNKQLLwFLdIdfLKnVIXENNVVILKVA
 zmet1 1156 Aa-----E---qaKLEvunin-----nLPVGEVefingPPCQIGSgnRRfn---qsswakvqcemILaFLafadfrfRRIELLENVtffvafn
 MET1 1168 AN-----E---laaKLiEcks-----tLPLqQVDfingPPCQIGSgnRRfn---qsswakvqcemILaFLafadfrfRRIELLENVtffvafn
 Motif VIII
 zmet2 569 dGVLCkVnISGLVAMtYQARZEMVAGCEVGLPFFNNTVtKALIEHmVLPtPLPTVAVRggaPnaEqcVAVdETQp-SLkKallIGDAISDLPKVqNhpNDVMEYG-QSP
 CMT1 468 KQFLARVAVASfVAMtYQARZEMVAGCEVGLPFFNNTVtKALIEHmVLPtPLPTVAVRggaPnaEqcVAVdETQp-SLkKallIGDAISDLPKVqNhpNDVMEYG-QSP
 zmet1 1236 KQdfrLAVASLLeGtQvRfGILeAGaFovqSskARfMApseGELPwEmviespEtkitipdgYyAaArStaGgAFRAItVdTGdPRVgngashtitEig-GeP
 MET1 1248 KQdfrLAVASLLeGtQvRfGILeAGaFovqSskARfMApseGELPwEmviespEtkitipdgYyAaArStaGgAFRAItVdTGdPRVgngashtitEig-GeP
 Motif IX
 zmet2 684 KTEFORVIRLRDnLwSfGeagdpdegKLIDHOPILANDDfERVQqIE-KGAMFRDLKGVVgqanNfEwDPeIerVLSGKpIVPQYANGsFIKHSILKPFGLWDEtVPT
 CMT1 385 KTEFORVIRLRDnLwSfGeagdpdegKLIDHOPILANDDfERVQqIE-KGAMFRDLKGVVgqanNfEwDPeIerVLSGKpIVPQYANGsFIKHSILKPFGLWDEtVPT
 zmet1 1332 VSWTQKkings-----mmvLndHIsKenNeInIRcQhIRKpGCDNDHdpceKvIsngqadEi-----ipwclPntArhrngwag---LYGKLdWeghNFT
 MET1 1364 VSWTQKkings-----tialDhICkANNeInIRcQhIRKpGCDNDHdpceKvIsngqadEi-----ipwclPntArhrngwag---LYGKLdWeghNFT
 Motif X
 zmet2 801 VVTRAEPRN--QVILHPEQRVtLRENNRUGQFFYRLGpIKERTIQGNVAVAVVAVtGCGAYlGeSGSDPLVQLPspFtsvggritagQARasvgtpagevveq
 CMT1 697 VVTRAEPRN--QcVILHPEQRVtLRENNRUGQFFYRLGpIKERTIQGNVAVAVVAVtGCGAYlGeSGSDPLVQLPspFtsvggritagQARasvgtpagevveq
 zmet1 1442 sVTdPQmgkvGmcfHPEQhRILLTVREcARsQGF7DStefagNiInhKIRIQGNVAVVpPLAVALGtKLKEAhlhKks---p-qhqp
 MET1 1454 sVTdPQmgkvGmcfHPEQhRILLTVREcARsQGF7DStefagNiInhKIRIQGNVAVVpPLAVALGtKLKEAhlhKks---p-qhqp

23/39
FIG. 9

	SAM binding		Cytosine binding	
Motif	<i>M.HhaI</i>	<i>zmet2a</i>	<i>M.HhaI</i>	<i>zmet2a</i>
I	Phe18	Try347		
II	Glu40	Gln407		
	Trp41	Trp408		
III	Asp60	Asp428		
IV	Pro80	Pro516	Phe79	Pro515
	Gln82	Gln82	Cys81	Cys517
V	Leu100	Val542		
VI			Glu119	Glu559
			Asn120	Asn560
			Val121	Val561
VIII			Arg165	Arg605
X	Asn304	Asn851		

24/39

FIG. 10

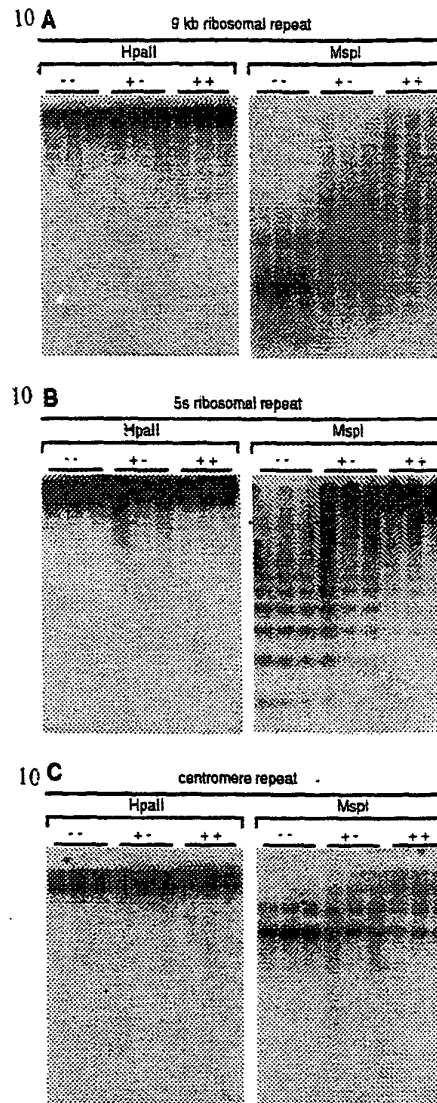
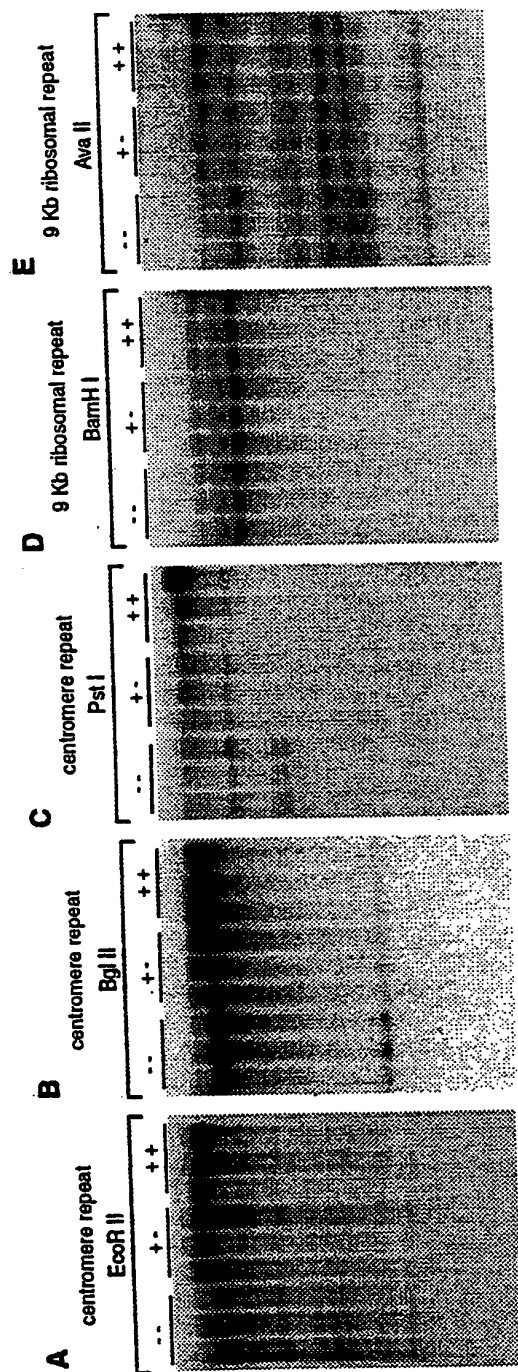


FIG. 11



26/39

FIG. 12

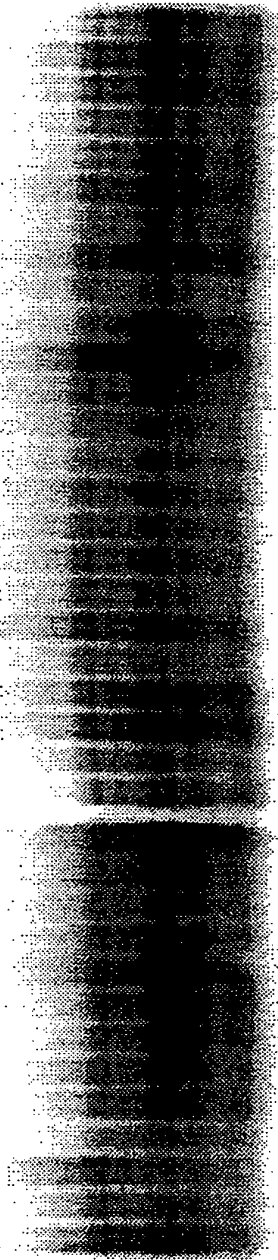
GENOTYPE	NUMBER OF PLANTS	TOTAL 5mCytosine (%)	% WT levels	% decrease
wild type	3	34.40 \pm 0.55	100	0.0
heterozygous zmet2a-mu1	7	32.00 \pm 0.90	93.0	7.0
homozygous zmet2a-mu1	5	30.40 \pm 0.19	88.4	11.6

2003-03-27 14:00:00

FIG. 13

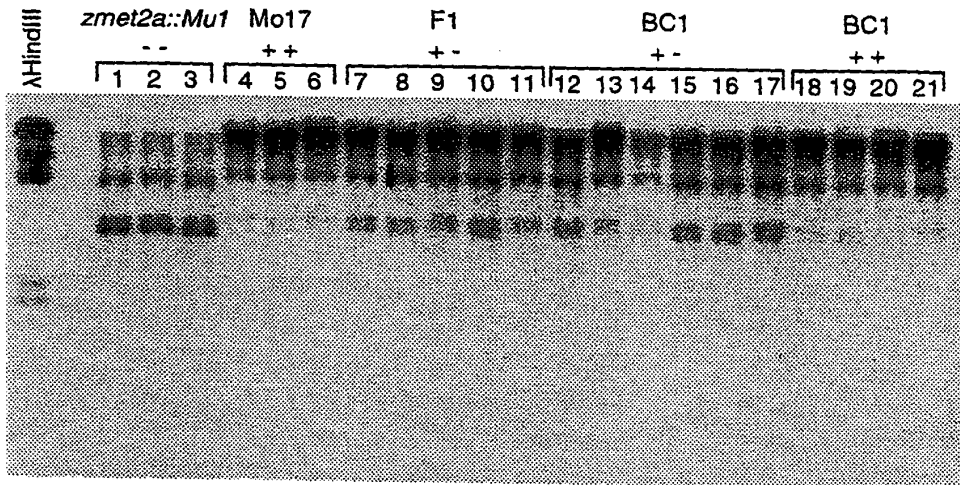
IIIH H V

B5	B5-4	B5-8	B4	B4-3	B4-9	B4-11
F3	F5	F5	F3	F4	F4	F4
	F6	F6		F5	F6	F5
						F6



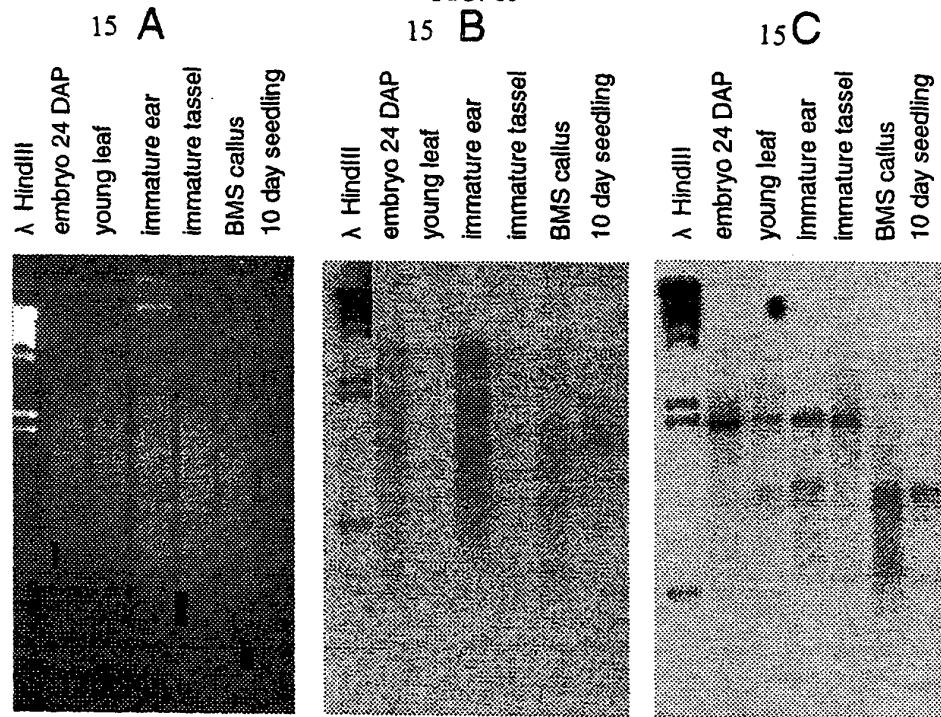
28/39

FIG. 14



29/39

FIG. 15



30/39

FIG. 16

5' LTR

catgc**TGTT**GGGCCCATGTGTCTAGTGTGGCCCATTAACGTGTACA
CATATACTAGAAGTGTGTGTGGTGTAGAGAGAGTGCTGTATGTTT
CCAC-TTCCAGAAAAATCC**ACAT**GGGTATCAGAGCCAGG

PBS

3' LTR

PPT

GAGGGGGAG**TGTT**GGGCCCATGTGTCTAGTGTGGCCCATTAACGTG
TACACATATACTAGAAGTGTGTGTGGTGTAGAGAGAGTGCTGTATG
TTTTCCACATTCCAGAAAAATCC**ACA**catgc

31/39
FIG. 17

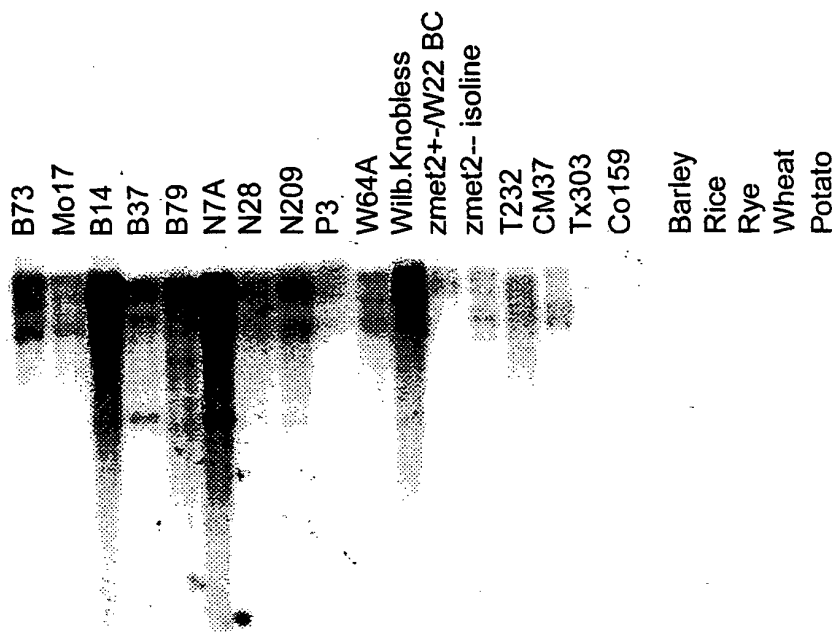
	Gag	Protease
SPRITE-1	- CYNCGNVGHIARNC	TQVTQLKWILDGASKH
hopscotch	- CQVCSRVGHTALNC	QNGSNVPWYDTGATDH
retrofit	- CQVCFKRGHTAADC	SYGIDTNWYIDTGATDH
arabpolprt-	CSNCGRTGHEKKEC	GKTKLGDIIIDSGASHH
copia	- CHHCGREGHIKKDC	SVMDNCGFVLDSGASDH

	Integrase
SPRITE-1	- QVKILRPDN-GTE YVNKGFNFLSRNGILHQTSCPDTPPONGVAERKNRHILE
hopscotch	- KIIAFQSDW-GGE--YEKLNAHFKTIGIHHQVSCPHTHQONGAAERKRRHIVE
retrofit	- KIIAMQTDWRGGR--YQKLSFFAQIGLIIMCHVLTILIRONGSAERKRRHIVE
arabpolprt-	TVKMVRSDN-GTE--FMCLSSYFRENGI IHQTSCVGTPOQNGRVERKRRHILN
copia	- KVVVLYIDN-GREYLSNEMRQFCVKKGI SYHLTVPHTPQLNGVSERMIRTI TE

	Reverse Transcriptase
SPRITE-1	- RYKARLVARGYSQTYGIDYDETFAPVAKMSTVTRLISCAANFGWPLYQLDVKNFLHGDQEEVYMEIPPG (59) AILAVYVDDIII
hopscotch	- RIKARLVAKGFKQQYGIDYDDTFSPVVKHSTIRLVLSLAVSQKWSLRQLDVQNAFLHGILEETVYMKQPPG (59) IYILVYVDDIII
Retrofit	- RYKARLVAKGFKQRYGIDYEDTFSPVVKAAITIRIILSIASRGWSLRQLDVQNAFLHGFLEEYVMQPPG (59) MFVLVYVDDIIV
Arabpolprt-	RYKARLVVQGNKQVEGEDYKETFAFVVRMTTVRTLRLNVAANQMEVYQMDVHNAFLHGDLEEYVMKLPFG (59) LRVLIYVDDLLI
copia	- RYKARLVARGFTQKQIDYEETFAFVARISSRFILSLVIQYNLKVHMDVKTAFLNGTLKEEYIMRLPQG (59) IYVLLYVDDVI

	RNase H
SPRITE-1	- DADWGSCLDDRRSTSGYCVFVGG-NLVSWRSKKQSVSRSTAEAEYRAMALAICEMLWIKGLL (25) NPVQHDRTKHVEIDRFF
hopscotch	- DADWAGCPDDRKSTGGYALFLGP-NLISWNSKKQSTVSRSTAEYKAMANATAEVIWLQSL (25) KPIFNARTKHIEVDHF
retrofit	- DADWAGSIDDRKSTGSFAVFLGS-NLVSWSARKQPTVSRSTAEYKAVANTTAEIIVQTL (25) NPVEFHARTKHIEVDYHF
arabpolprt-	DSDWQSCPLTRRSISAYVVLGG-SPISWKTKKQPTVSHSSAEAEYRAMSYALKEIKWLRKLL (25) NPVEFHRTKHIESDCHS
copia	- DSDWAGSEIDRKSTTGYLEKMFENLWCNTRQNSVAASSTAEYMALFEACREALWLKFL (25) NPSCHKRAKHIDIKYHF

FIG. 18



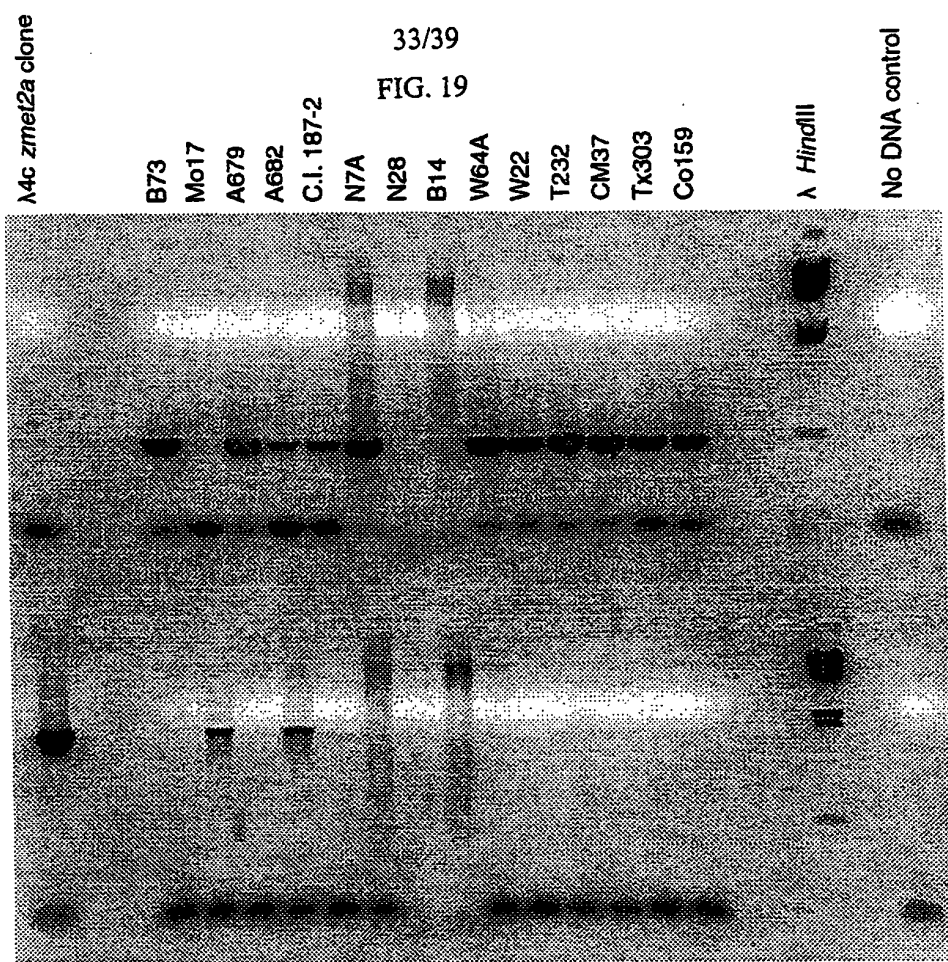
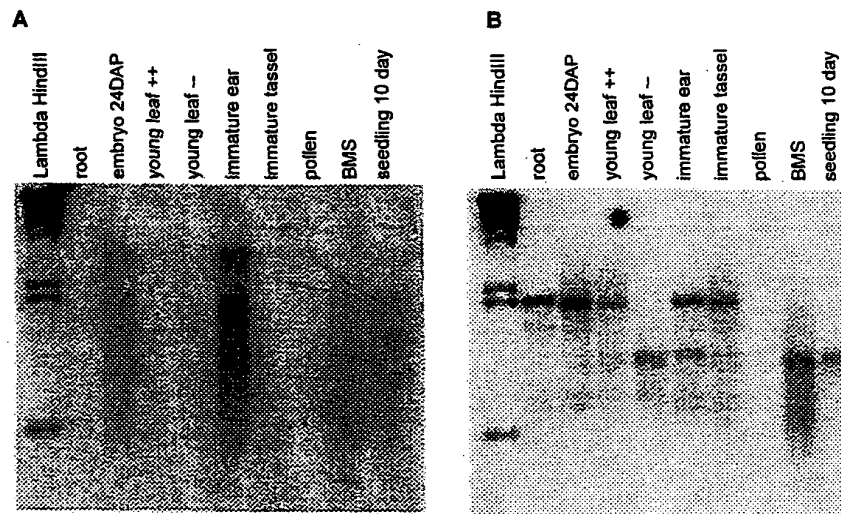
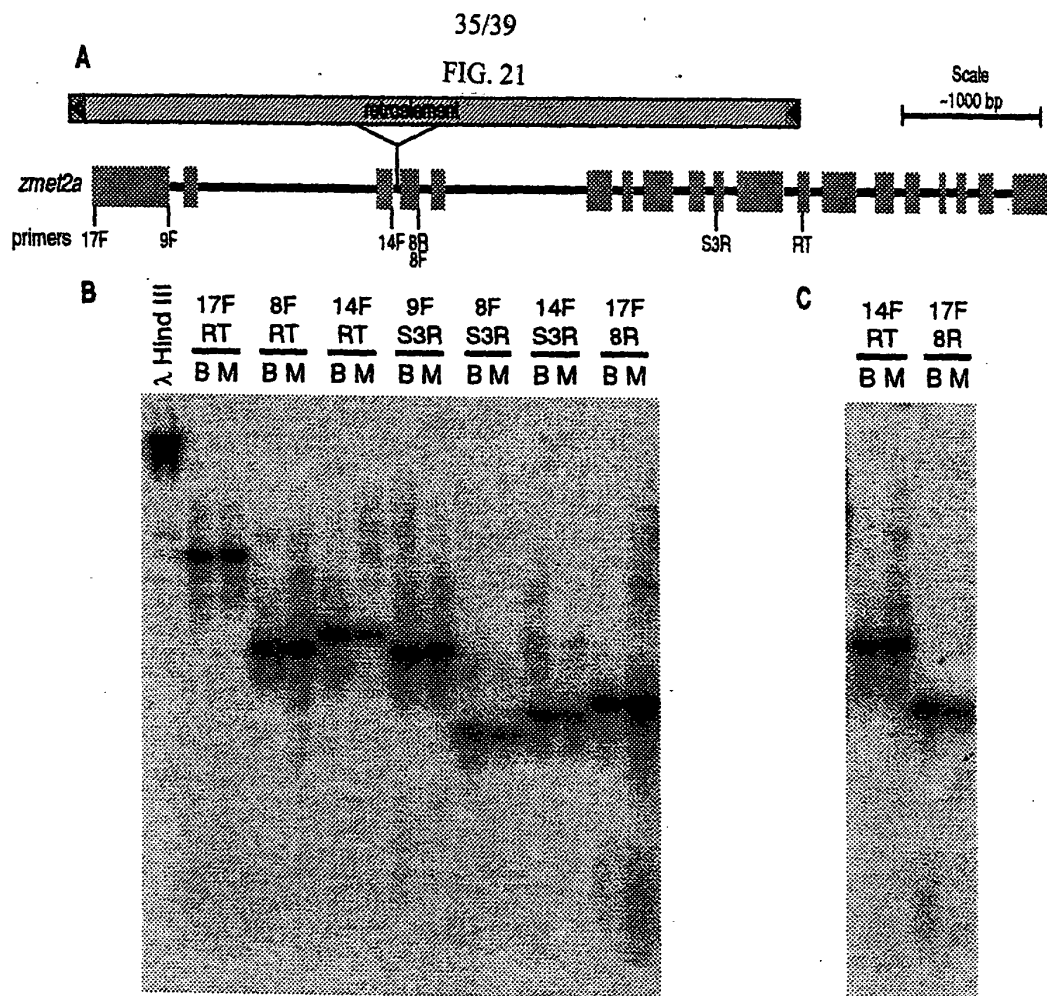


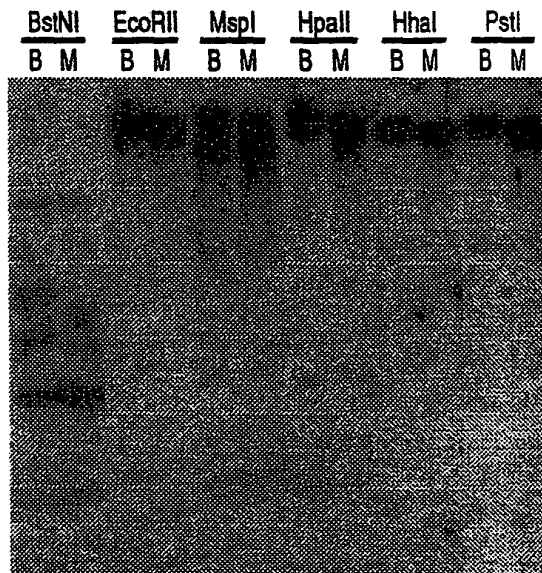
FIG. 20





36/39

FIG. 22



37/39

FIG. 23

GGGAATTCGATTACTCACTATAGCGCTCGAGCGGCCGCCGGGCAGGTTGAAAAACCATC
AACCTAACGATGTAATGGAGTATGGTGGTTCCCCAAGACAGAGTTCCAGCGCTACATTC
GACTTGGTCGTAAAGACATGTTGGATTGGTCGTTTGGTGAGGAGGCTGGTCCAGATGAAG
GCAAGCTCTTGGATCACCAGCCCTTACGGCTTAACAATGATGATTATGAGCGGGTTAAGC
AAATTCTGTCAAGAAGGGAGCCAACTTCCGTGACCTAAAGGGTGTCAAGGTTGGAGCAA
ATAATGTTGTTGAGTGGGATCCAGAAAGTCGAACGTGTGTACCTTTCGTCTGGGAAACCAC
TGGTTCCTGACTATGCGATGTCATTCATCAAGGGCAAATCACTCAAGCCATTCTGGGCGCC
TGTGGTGGGACCAGACGGTTCTACAGTTGTGACCAGAGCAGAGCCTCATAACCAGGTTAT
ATTGCATCCGACTCAAGCAAGAGTCTTGACTATCCGGGAGAACGCAAGGTTACAGGGCTT
CCCCGATTACTACCGATTGTTTGGACCGATCAAGGAGAAGTATATTCAAGTCGGGAACGC
AGTGGCAGTCCCTGTTGCACGGGCACTGGGCTACTGTCTGGGTCAAGCCTACCTGGGTGA
ATCTGACGGGAGTCAGCCTCTGTACCAGCTGCCTGCAAGTTTTACCTCTGTGGGGCGAAC
CGCGGTTCAAGGCGAATGCCGCTTCTGTTGGCACTCCTGCGGGGGAGGTAGTCGAGCAGTA
AAAGGATAGCGGAGCAACCCTGGTTGGTATTTTGATTGAGCCCATCCAGTAGCATGTTT
ACCAATAAATAATCATTGGTCGTGCTGATTCTTATGGTTGGAGATGAATGTATGTAGGGT
GTACTCGAGCTCGAGTGCTTGTGTTGACTGTAGGTTGAGGTTTCTCATCCATTGGCCTGCC
TATTTGTGGATGACGTTTCATTTTCAGATTAGCAATGTGCTTATTTAAGGTTTCGTCATGT
ACCTGTATTCTACAATCCACTATTGTTTCCAAAGACAGCATTGATCCTTAAAAAAAACCT
GTAAAAAAAACAGTGCCCGAAAAGCCGCAAAAAAAAACCTGCC
GGGCGGCCGCTCGAGCCCTATAGTGAGTAATCGAATTCCC

FIG. 24

[illegible]

39/39

FIG. 25

667 KVQNHQPNDVMEYGGSPKTEFQRYIRLSRKDMLDWSFGEGAGPDEGKLLDHQPLRLNDD 726
+ +NHQPNDVMEYGGSPKTEFQRYIRL RKDMLDWSFGE AGPDEGKLLDHQPLRLNDD
15 RFENHQPNDVMEYGGSPKTEFQRYIRLGRKDMLDWSFGEEAGPDEGKLLDHQPLRLNDD 74

727 YERVQQIPVKKGANFRDLKGVVRVGANNIVEWDPEIERVKLSSGKPLVPDYAMSFYKGS 786
YERV+QIPVKKGANFRDLKGV+VGANN+VEWDPE+ERV LSSGKPLVPDYAMSFYKGS
75 YERVQIPVKKGANFRDLKGVKVGANNVVEWDPEVERVYLSSGKPLVPDYAMSFYKGS 134

787 KPFGRLWWDQTVPTVVTRAEPHNQVIHPTQARVLTIRENARLQGFPDYRLFGPIKEY 846
KPFGRLWWD+TVPTVVTRAEPHNQVI+HPTQARVLTIRENARLQGFPDYRLFGPIKEY
135 KPFGRLWWDQTVPTVVTRAEPHNQVILHPTQARVLTIRENARLQGFPDYRLFGPIKEY 194

847 IQVGNAVAVPVARALGYCLGQAYLGESESDPLYQLPPSFTSVGGRTAGQARASPVGTPA 906
IQVGNAVAVPVARALGYCLGQAYLGES+GS PLYQLP SFTSV GRTA QA A+ VGTPA
195 IQVGNAVAVPVARALGYCLGQAYLGESDGSQPLYQLPASFTSV-GRTAVQANAASVGTPA 253

907 GEVVEQ 912
GEVVEQ
254 GEVVEQ 259